Slide 1 of 12

Title slide: Acquiring Micro-level NCES Data

Slide 2 of 12

This module introduces users to the NCES micro-level datasets and explains how to access them. It explains the differences between public-use and restricted-use data and describes the processes for acquiring public-use datasets and their supporting materials. Additionally, this module explains what constitutes restricted-use data, describes the reasons for NCES having restricted-use data, and provides information related to acquiring restricted-use datasets and their supporting materials.

Slide 3 of 12

Micro-level data are detailed data about each individual survey respondent. In a microdata file, each respondent's record includes his or her responses to each survey item and any additional constructed variables and/or data added from external sources. An example of data from an external source would be data about the school in which the student is enrolled that are taken from the NCES school universe data.

It is important to note that microdata files can include data about either individuals or institutions. In studies there is often one set of microdata for students, one set for teachers of the students in the study, and one set for the school the student attends.

Slide 4 of 12

NCES is required by law to protect the confidentiality of respondents to NCES surveys who provide individually identifiable data. Whenever possible, NCES applies disclosure protections to micro-level data to allow for the release of public-use data. However, there are times when researchers need to use the more detailed restricted-use data to conduct a specific analysis. There are also instances where the amount of data perturbations and protections that would be required to produce a public-use file are deemed to make the data unusable for research and analytic purposes, so only a restricted-use file is produced. NCES established a licensing system to provide qualified researchers access to restricted-use files, which will be discussed later.

The first step in protecting the confidentiality of individual respondents is to remove all direct identifiers, such as names or addresses from micro-level data files. NCES then uses perturbation techniques to introduce a controlled amount of noise into each data file. These perturbations are used to guard against a data user matching a micro-level data file against an external micro-level data file and identifying individual respondents with certainty through such matches. These perturbations and protections are included in both restricted-use and public-use data files and do not affect the overall data quality.

To produce a public-use data file, NCES staff conducts disclosure risk analyses to identify data elements that, if released as reported, could disclose the identity of an individual survey respondent. Data elements that are identified as disclosure risks are masked or coarsened by recoding using top coding, bottom coding, or categorical recoding throughout. In some cases a data element is suppressed from the public-use file completely.

Slide 5 of 12

NCES produces codebooks, ASCII data files, PDF versions of the survey User Manual, and PDF versions of the data collection instruments, with the exception of material that is copyright-protected.

There are a couple of different ways to acquire public-use datasets.

Some NCES data collections provide Electronic Codebook, or ECB, software on a CD-ROM or DVD to facilitate the use of NCES data.

Other NCES data collections make the data available for download in various formats and software programs. Some NCES data collections make datasets available both ways. Specific instructions for acquiring individual datasets are included in the survey-specific modules.

Slide 6 of 12

Restricted data includes information that could enable the identification of specific schools, individual administrators, teachers, or students; therefore it is masked or suppressed in a public-use file. Such information is confidential and protected by laws that provide for the protection of individually identifiable data. Failure to comply with these laws is punishable as a Class E felony.

Restricted-use data files contain more detailed information than the public-use files. Thus restricted-use data files may be considered sensitive or unique. Researchers who are interested in accessing restricted-use data must apply for, and be granted, a restricted-use data license from NCES.

NCES is required by law to protect the individually identifiable data provided by NCES survey respondents under a pledge of confidentiality. Under the applicable confidentiality laws, individuals granted access to restricted-use data are not permitted to publish any data that could be used to identify an individual respondent. They are also prohibited from using the data for a purpose other than the statistical purpose for which the data were supplied. Individuals granted access to restricted-use data must provide a sworn statement agreeing to observe the limitations of the law. Anyone granted access who knowingly publishes or communicates any individually identifiable information will be subject to fines of up to $250,000 and/or up to 5 years in prison.

Slide 7 of 12

The restricted-use license is a contract through which IES/NCES loans a qualified organization the data under certain terms and conditions. The requestor must state a justification for the data use, which should demonstrate both that the requestor's proposed use is statistical and the research question can be answered with data available in the restricted-use file being requested and that the research question cannot be answered adequately with data available in a public-use file. NCES does not evaluate the merits of the proposed line of research, just that the line of research can only be addressed with restricted-use data.

Additionally, the requestor must submit the required documents, agree to keep the data secure from unauthorized disclosure at all times, agree to unscheduled inspections of compliance by the IES data security personnel to ensure compliance, agree to not publish or release unrounded, unweighted counts, and agree to submit all work products to NCES for disclosure review prior to sharing with anyone not authorized on a license.

While this training has referred only to a license up to this point, there are actually two types of permissions for acquiring restricted-use data: a restricted-use license, which is granted to universities and other non-governmental organizations, and a memorandum of understanding, which is an agreement entered into with other federal agencies.

Slide 8 of 12

The researcher needs to determine several parties to the license. The Principal Project Officer, or PPO, is responsible for day-to-day operations involving the requested data. The PPO must be at the level of a Postdoctoral researcher or higher or he or she must be an experienced researcher in an independent research organization. The Senior Official, or SO, is a person in the organization with the legal authority to sign a contract for the organization. The System Security Officer, or SSO, oversees security of the data. A license may also include up to seven authorized users who are granted permission to access the data through the license.

Researchers below the post-doctoral level can access restricted-use data by being listed as an authorized user on someone else's license. Each authorized user must sign a notarized affidavit of nondisclosure that names the data sets that user has received approval to access. The Principal Project Officer and Senior Official must understand and acknowledge that they are responsible for overseeing the use of the data by anyone listed as an authorized user on their license, and that they will be held accountable for any misuse.

Slide 9 of 12

Documentation for a License Application starts with a formal request which is filled out online via the web. To request a new license, click on Getting Started at http://nces.ed.gov/statprog/instruct.asp. Then click on either the button that says “I need a restricted-use data license” or “I need a Memorandum of Understanding.”

Slide 10 of 12

An IES License Document and a security plan must be submitted. In addition, affidavits of nondisclosure will need to be signed, notarized, and submitted for each authorized restricted-use data user named in your license documents. Each authorized researcher must read the Restricted-Use Data Procedures Manual. If you anticipate requesting additional years or waves of data from one NCES study, you should note first year of data you are requesting and “all future years” on each affidavit of nondisclosure. You will be required to submit new signed and notarized affidavits of nondisclosure with each new request for additional data.

The online system will ask for: the contact information of the Principal Project Officer, Senior Official, and System Security Officer; the name of the dataset requested; the purpose for the loan of the data, which includes your justification for needing restricted-use data; and the length of time the data are needed.

The License Document, Affidavit, and Security Plan Form should be mailed to the address shown on the slide. If you have any questions about the licensing process, contact NCES via email at IESData.Security@ed.gov.

Slide 11 of 12

Once you have obtained a restricted-use data license, you may also use the online licensing system to request an amendment to an existing license. Amendments include requests to add or remove researchers from your license or requests to add additional data to your license. You will need to provide a description of your research and the justification for needing additional restricted-use data. You will also need to enter the license number and Principal Project Officer’s email address.

Slide 12 of 12

This module has provided you with information needed to assist you in accessing NCES micro-level datasets. Important resources that have been provided throughout the module are summarized here along with the module’s objectives for your reference.

Users may elect to assess their learning by answering multiple choice questions pertaining to this module through the Check Your Knowledge option located at the bottom of the screen. This is an optional tool that users may use at any time to review key concepts. You may now Check Your Knowledge, proceed to the next module in the series, or click the exit button to return to the landing page.